



| Teaching Guide | | | | |
|---------------------|--|--------|---|---------|
| Identifying Data | | | | 2021/22 |
| Subject (*) | Information Retrieval | Code | 614G02027 | |
| Study programme | Grao en Ciencia e Enxeñaría de Datos | | | |
| Descriptors | | | | |
| Cycle | Period | Year | Type | Credits |
| Graduate | 2nd four-month period | Third | Obligatory | 6 |
| Language | Spanish | | | |
| Teaching method | Face-to-face | | | |
| Prerequisites | | | | |
| Department | Ciencias da Computación e Tecnoloxías da Información | | | |
| Coordinador | Parapar López, Javier | E-mail | javier.parapar@udc.es | |
| Lecturers | Hasan Romero, Ismael Landín Piñeiro, Alfonso Parapar López, Javier | E-mail | ismael.hasan@udc.es alfonso.landin@udc.es javier.parapar@udc.es | |
| Web | www.dc.fi.udc.es/~parapar/ | | | |
| General description | <p>Traditionally, information retrieval systems have been used by documentalists, librarians or lawyers to search for records. Today the situation has changed radically, hundreds of millions of people use information retrieval systems on a daily basis: they search the web, they search their mailbox, they search inside their computer or they receive recommendations for content consumption. Information retrieval has become the dominant area of information access overtaking traditional databases. Information retrieval systems are capable of solving user needs on unstructured text without the need for users to specify their query in a standard way. This course will explore the theoretical concepts that support information retrieval and access systems, as well as the software and tools for the construction of advanced search and filtering systems.</p> | | | |
| Contingency plan | <p>1. Modificacións nos contidos Sin modificacións. 2. Metodoloxías *Metodoloxías docentes que se manteñen Todas *Metodoloxías docentes que se modifican Ninguna 3. Mecanismos de atención personalizada ao alumnado Mail, teams e moodle. 4. Modificacións na avaliación Ningunha *Observacións de avaliación: Esta asignatura está planeada como híbrida xa que a proba final realizarase de forma presencial si a situación o permite. 5. Modificacións da bibliografía ou webgrafía Sin modificacións</p> | | | |

| Study programme competences / results | |
|---------------------------------------|--|
| Code | Study programme competences / results |
| A27 | CE27 - Compresión e dominio de fundamentos e técnicas básicas para a procura e o filtrado de información en grandes coleccións de datos. |
| B2 | CB2 - Que os estudantes saiban aplicar os seus coñecementos ao seu traballo ou vocación dunha forma profesional e posúan as competencias que adoitan demostrarse por medio da elaboración e defensa de argumentos e a resolución de problemas dentro da súa área de estudo |
| B3 | CB3 - Que os estudantes teñan a capacidade de reunir e interpretar datos relevantes (normalmente dentro da súa área de estudo) para emitir xuízos que inclúan unha reflexión sobre temas relevantes de índole social, científica ou ética |
| B4 | CB4 - Que os estudantes poidan transmitir información, ideas, problemas e solucións a un público tanto especializado como non especializado |
| B7 | CG2 - Elaborar adecuadamente e con certa orixinalidade composicións escritas ou argumentos motivados, redactar plans, proxectos de traballo, artigos científicos e formular hipóteses razoables. |
| B8 | CG3 - Ser capaz de manter e estender formulacións teóricas fundadas para permitir a introdución e explotación de tecnoloxías novas e avanzadas no campo. |
| B9 | CG4 - Capacidade para abordar con éxito todas as etapas dun proxecto de datos: exploración previa dos datos, preprocesado, análise, visualización e comunicación de resultados. |
| B10 | CG5 - Ser capaz de traballar en equipo, especialmente de carácter multidisciplinar, e ser hábiles na xestión do tempo, persoas e toma de decisións. |
| C1 | CT1 - Utilizar as ferramentas básicas das tecnoloxías da información e as comunicacións (TIC) necesarias para o exercicio da súa profesión e para a aprendizaxe ao longo da súa vida. |



| | |
|----|---|
| C4 | CT4 - Valorar a importancia que ten a investigación, a innovación e o desenvolvemento tecnolóxico no avance socioeconómico e cultural da sociedade. |
|----|---|

| Learning outcomes | | | |
|--|---------------------------------------|-----------------------|----------|
| Learning outcomes | Study programme competences / results | | |
| To know, understand and analyze the different Information Retrieval models, the techniques for their efficient implementation and their evaluation methodology. | A27 | B3 B4 | C1 C4 |
| To know, understand and analyze the software platforms for the creation of these systems. | A27 | B2 B4 B9 B10 | |
| Plan and perform the evaluation of Information Retrieval systems. Analyze the results of the evaluation of IR systems to improve their effectiveness and efficiency. | | B7 B8 | C1 C4 |
| To be able to correctly deal with the ethical, privacy, confidentiality and security aspects of these systems. | A27 | B4 B9 | C4 |

| Contents | |
|--|---|
| Topic | Sub-topic |
| Basic Search Engine Architecture | The basic architecture of a search engine |
| Text Analysis and Processing | From the document to the index tokens |
| Inverted Index and Query processing | Inverted files and query processing strategies |
| Classic Information Retrieval Evaluation | Offline evaluation |
| Modern Information Retrieval Evaluation | Beyond offline evaluation: online evaluation, user satisfaction |
| Boolean and Vector Space Models | Basic retrieval models |
| Language Models | Statistical language models |
| Feedback and Query Operations | Relevance feedback and query reformulation |
| Link Analysis | Web graph analysis |

| Planning | | | | |
|---------------------------------|------------------------|--------------------------------------|-------------------------------|-------------|
| Methodologies / tests | Competencies / Results | Teaching hours (in-person & virtual) | Student?s personal work hours | Total hours |
| Laboratory practice | B2 B7 B9 B10 C1 | 14 | 42 | 56 |
| Supervised projects | B4 B7 B9 | 5 | 7.5 | 12.5 |
| Mixed objective/subjective test | A27 B2 B4 B7 B8 | 2 | 13 | 15 |
| Guest lecture / keynote speech | A27 B3 B4 B8 C4 | 19 | 47.5 | 66.5 |
| Personalized attention | | 0 | | 0 |

(*)The information in the planning table is for guidance only and does not take into account the heterogeneity of the students.

| Methodologies | |
|---------------------------------|---|
| Methodologies | Description |
| Laboratory practice | Practical assignments on development platforms widely used in the industry, search engine companies and research libraries. |
| Supervised projects | Work and problems carried out autonomously by the student and supervised by the teacher. |
| Mixed objective/subjective test | Test that will focus on the fundamental contents of the course. |



| | |
|--------------------------------|---|
| Guest lecture / keynote speech | The student will attend the explanations of the professor about the different models, techniques and algorithms of Information Retrieval. The professor will use different levels of abstraction-detail and will guide the student in the fundamental and complementary readings. |
|--------------------------------|---|

Personalized attention

| Methodologies | Description |
|--|--|
| Laboratory practice Supervised projects | Laboratory work and tutored work: In addition to evaluating the result of the practice in accordance with the requirements, the development of the same is monitored. The student's autonomy must be respected so that he/she acquires greater skills with the software platforms used, but the teacher will be able to solve certain difficulties that may block the student an excessive time given the planning of the subject. |

Assessment

| Methodologies | Competencies / Results | Description | Qualification |
|---------------------------------|------------------------|--|---------------|
| Laboratory practice | B2 B7 B9 B10 C1 | Follow-up, defense and evaluation of the results of the practices carried out in the hours of practical laboratory classes. It is mandatory to achieve 40% of the grade in order to pass the course. 40 | 40 |
| Supervised projects | B4 B7 B9 | Participation and results in the completion of the work and/or questions. | 10 |
| Mixed objective/subjective test | A27 B2 B4 B7 B8 | Questions on the knowledge acquired in the lectures, practical activities and problems and assignments. It is mandatory to achieve 40% of the grade to pass the course. | 50 |

Assessment comments

For the second opportunity and non-ordinary exams, the mixed exam will evaluate the practices as well as the papers and the theories. If the minimum grade is not reached in the different tests, the student's maximum grade will be 4.5. In the realization of the work, plagiarism and the use of non-original material, including that obtained through the Internet, without express indication of its origin and, where appropriate, permission of its author, may be considered as a reason for a failing grade. All this without prejudice to the disciplinary responsibilities that may arise after the corresponding process.

Sources of information

| | |
|----------------------|--|
| Basic | <ul style="list-style-type: none"> - W.B. Croft, D. Metzler, T. Strohman (2009). Search Engines. Information Retrieval in Practice. Pearson Education - C.D. Manning, P. Raghavan, H. Schütze (2008). Introduction to Information Retrieval. Cambridge University Press - Baeza-Yates and B. Ribeiro-Neto (2011). Modern Information Retrieval (second edition). Addison Wesley/Pearson Education - F. Casheda, J.M. Fernández, J. Huete (editores) (2011). Recuperación de Información. Un enfoque práctico y multidisciplinar. Ra-Ma |
| Complementary | <ul style="list-style-type: none"> - Ian H. Witten (1999). Managing Gigabytes: Compressing and Indexing Documents and Images. Morgan Kaufmann - Amy N. Langville, Carl D. D. Meyer (2011). Google's PageRank and Beyond: The Science of Search Engine Rankings. Princeton University Press |

Recommendations

Subjects that it is recommended to have taken before

Subjects that are recommended to be taken simultaneously



| |
|-------------------------------------|
| Subjects that continue the syllabus |
| |
| Other comments |
| |

(*)The teaching guide is the document in which the URV publishes the information about all its courses. It is a public document and cannot be modified. Only in exceptional cases can it be revised by the competent agent or duly revised so that it is in line with current legislation.